EARLY DETECTION OF TUBERCULOSIS USING HYBRID ADABOOST AND LEARNING VECTOR QUANTIZATION TECHNIQUES

A SYNOPSIS

Submitted by

EMIL.M. PAUL

(Reg. No. 201708252)

in partial fulfillment for the award of the degree

of

DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

SCHOOL OF ELECTRONICS ELECTRICAL AND BIOMEDICAL TECHNOLOGY

KALASALINGAM ACADEMY OF RESEARCH AND EDUCATION

(DEEMED TO BE UNIVERSITY)

ANAND NAGAR, KRISHNANKOIL-626126

TAMILNADU, INDIA.

DECEMBER 2022

1. INTRODUCTION

The chest X-ray films gives certain vital information about the patients. However, the researchers are been encouraged to develop computer algorithms to assist the radiologist in diagnosis process. Automatic segmentation is one of the first steps of such computer-aided systems. The screening technique is extended with the help of creating the reviewing procedure faster via reducing eye fatigue. In this paper we tend to discuss, regarding the varied automatic approaches for sleuthing the TB. Medical imaging technological know-how could be facts managing an integrated system designed for the screening of T.B. The Internet will trade in medical imaging techniques and digital image applications with extensive changes in connectivity and technology. Digital images play a nominal role in today's communication. The TB is caused by the bacteria that transmits through the air. The X-ray chest image of the infected person is mandatory part for TB evaluation. The major sources of noises from this X-ray images are filtered in the pre-processing stage[1].The most common noises are suppressed by the filters such as median filter, wiener filter and other morphological filters(for image enhancement)[2].The pre-processed image is then fed for segmentation using marker based segmentation after it is been compared with various segmentation methods(ROI)[3].The performance analysis made through precision and recall parameters. The feature extraction of the chest X-ray images is depicted in the paper[4]. The various features are then taken for comparison LBP, HOG, Tamura features, Texture features, GLCM, shape features. The sub features some included are coarseness, contrast, directionality etc in tamura features.[5].The GLCM features includes various features like energy, homogeneity, contrast etc. The various shape features for the ROI of the chest X-ray images such as it's eccentricity, solidity, convexity, rectangularity ,elliptic variance are depicted in the paper[6].The speeded up Robust features are listed in the paper. The classification methods such as SVM, Neural and Adaboost are classified and illustrated detailly in the papers. The performance parameter of a classifier can be analysed with various parameters like accuracy, precision, recall, and F1 score. Section II explains the literature review, Section III explains the objectives, and then the research problem is identified and the result of the proposed work is explained in Section VII. This proposed work is concluded with Performance analysis.

2. LITERATURE REVIEW

A patient who is affected by HIV is more prone to TB [7] . Before the advancement of technology, several tests such as skin tests, molecular tests were done. But these detection methods yield poor results. So, detection became a tedious task [8]. There are several antibiotics available in the medical stores for the cure of TB. But the challenge is to identify the active TB patient accurately [9]. By the intervention of technology, an automated approach has been developed. So, the posteroanterior radiograph of the patient's chest has become a mandatory part of the TB evaluation. The Roentgenogram (X-ray) is a plain film. When radiation passes through the person, it casts a shadow on the film. The unexposed film is white, whereas a fully exposed film is black in colour[10]. As the X-ray has low absorption coefficients, it can penetrate through the bones, giving doctors easy access to read the internal part of a human being by simply looking into the X-ray[11]. However, the research paved a way to analyse the computer-aided diagnosis for TB detection with the help of algorithms [12]. In our work, it is observed that the X-ray image of the patient who is suspected of having TB is taken as input. The input image is then fed for pre-processing, segmentation techniques, feature extraction methods, and classification [13]. Our work combines the algorithm of the marker-based watershed method along with the combination of normalization of the input image, gamma correction, and k-means clustering [14]. This enhances the Region of Interest (ROI) of the segmented portion[15]. The ROI is then fed to various feature extraction methods such as Speeded Up Robust Features (SURF) and Principal Component Analysis (PCA)[16].The SURF uses Hessian Matrix[17].The extracted features are then compared with the trained feature set in the database, then fed to the classifier[18].The classifier includes SVM,Neural,Adaboost and LVQ[19]. The classifier here we used is AdaBoost combined with Learning Vector Quantization (LVQ), and the fused output is combined using weights along with the outputs from both the classifier [20]. This is the most efficient way for classification[21].The base leaners of Adaboost determines the weighted function[22]. The accuracy yield to be 98.52%, after its comparison with Support Vector Machine (SVM) and Neural networks[23].The SVM uses kernel function,either polynomial kernel[24].The Time of Execution,precision,recall values,F1 score all works well with this hybrid combination [25].The review of literature is explained in Section II. The proposed method is detailed in Section III, preceded by the model’s outputs, which are described in Section IV.

3. OBJECTIVES OF THE WORK

The main objectives of this research work are

1. Implementing Top hat and Bot hat filters for image enhancement after pre processing.
2. To develop a merged segmentation method using Marker Based Segmentation method and K means Clustering.
3. To develop a novel image processing algorithm using Adaboost and Learning Vector Quantisation and evaluate it with standard performance measures.
4. To implement the concept of merging Adaboost and LVQ to yield the better result while comparing it with other algorithms that are employed.

4. RESEARCH PROBLEM

The fastest evolution of the digital image processing led to the development of the finest ways in the detection of various diseases in the field of medical images. The problem arises due to unreliable results, yields to the death of the person. The major challenge here lies in the proper diagnosis, for which we used reliable algorithms. Choosing an algorithm in such a way that we can accomplish quality aspects like greater privacy, secrecy, integrity, and efficiency. The proposed study uses a suitable way to diagnose the Tuberculosis, which includes the hybrid combination of the Adaboost and Learning Vector Quantization. The training is done in such a manner that the error data obtained undergo several iterations, by applying maximum weight to the error sample compared to the correct sample. The accuracy, sensitivity, recall, F1 parameters proved that they work fine with the Hybrid combination of Adaboost and LVQ when compared to SVM, and other neural networks.

5. CONTRIBUTIONS

In this work, we have detected TB reliably from the Chest X-ray images using image pre-processing, merged segmentation methods which is then followed by classification of Adaboost and LVQ methods. This developed method yields the better and reliable results for precision, recall and F1 score. Several Public data bases were used to create a database of 2500 normal and infected X-ray chest images for this study. The merged combination of marker-based watershed segmentation with K means clustering is used to obtain the ROI after the image normalization and

gamma correction. Here the value of K is kept as 3 ,2 for both the lungs and 1 for the background. In this chest Xray image normalisation ,it is the process which changes the range of pixel intensity values. Sometimes normalization is called as contrast stretching or histogram stretching. The resulted ROI is used for feature extraction. The various features like entropy, skewness, contrast, correlation etc. and many shape features like area, major axis, minor axis ,convexity etc are found better using SURF and PCA. The speed of execution is faster when compared to other feature extraction methods. The box and Hessian matrix utilization in SURF, makes it more faster .It makes use of Differentiation of Gaussian on integral Images. The classification is done by the fused hybrid combination of Adaboost and LVQ.

6. ORGANIZATION OF THE THESIS

The research work is organized into four chapters and is as follows:

Chapter 1

This chapter covers median filter ,wiener and image enhancement techniques for the pre processing. .Median filter has a non-linear approach that works great with impulse noise (salt & pepper noise) and speckle noise while preserving sharp edges in the image .They describe and analyze an algorithm for cleaning speckle noise in medical images. The Weiner filter used reduces MSE to a least value. The top hat and bot hat filters are used for image enhancement. A lower value for MSE means error is fewer and a higher value of PSNR is preferred. Here, the 'signal' is the original image, and the 'noise' is the error in reconstruction. So, a filtering method having a lower MSE and a high PSNR can be recognized as a better one.

Chapter 2

In this chapter the Input chest Xray image is normalised by varying the gray scale pixel value to a meaningful value. The gamma correction that bring out the brightness of the gray scale image. It is then fed to the K means clustering method .The value of K is determined as 3,(2 for lungs and 1 for background).The pre-processed image from the input is fed to the marker based watershed segmentation. The output of both the segmentation methods yields the desired ROI, which is then fed to the next section for feature extraction. The Tuberculosis (TB)

can be fastly and accuratelyidentifiable by the merged segmentation methods of K means with Marker-based Watershedsegmentation which has highest precision and recall values when compared to the several other segmentation methods which is been discussed.

Chapter 3

This chapter explains about the feature extraction methods used. Let us take feature 1 as the combination of HOG, LBP, Tamura features, Texture Features, Shape features and GLCM and feature 2 as the combination of SURF,HOG,PCA. The performance parameter of both the features are being checked out between the number of runs and the time elapsed for the number of runs and a graph is plotted between the number of runs requires for each feature set for the elapsed time duration in seconds, which ensure that feature 2 excels feature 1.

Chapter 4

This chapter covers , a fastened technique with the merged combination of AdaBoost and Learning Vector Quantization for determining TB in an easier way with the input chest X-ray image of a person with the aid of computer-aided diagnosis with greatest accuracy, precision, recall and F1 values. This finest technique got an accuracy of 98.52%when compared to the prior conventional methods used such as SVM, Convolutional Neural Network, etc Compared to the same parameters with other existing algorithms, the HE algorithm efficiently estimated PSNR and MSE.

7. OVERVIEW OF SEGMENTED RESULTS

A lower value for MSE means error is fewer and a higher value of PSNR is preferred. Here, the 'signal' is the original image, and the 'noise' is the error in reconstruction. So, a filtering method having a lower MSE and a high PSNR can be recognized as a better one. The calculated value of PSNR and MSE for both mean and proposed filters are shown in Table 1.

Table 1.Shows the comparison between the Mean filter and the Proposed filter

MEAN FILTER	PSNR-19.340	MSE-15.5603
PROPOSED FILTER	PSNR-25.5913	MSE-13.3961

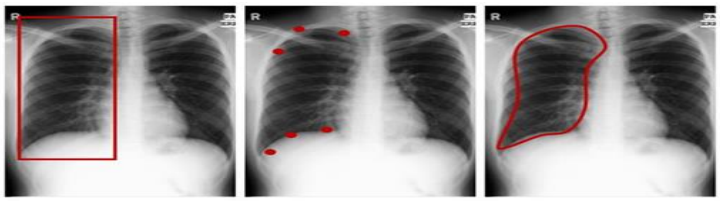

Figure 1. ROI (Region Of Interest) after Markerbased Segmentation

The Figure 1 shows the region of interest of the Marker based watershed segmnentation and K Means clustering.

Table 2. Represents the precision values for various segmentation methods used

Input Image		ROI Segmentation	Watershed	Otsu	Region Growing	Marker Based Watershed	K -means Clustering
1			.575	.220	.015	.935	.864
2			.560	.222	.069	.940	.905
3			.537	.244	.055	.854	.899
4			.587	.287	.038	.876	.919
5			.571	.235	.020	.899	.920
6			.525	.285	.025	.915	.932
7			.515	.294	.064	.924	.949

The Table 2 represents the precision values of the different segmentation methods like otsu,region growing,marker based watershed segmentation and K means

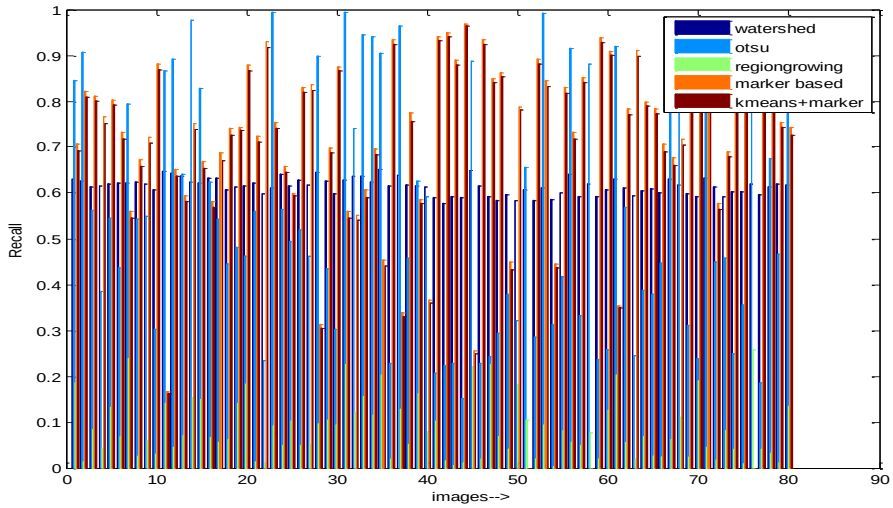

Figure 2. Represents the comparion between the Recall values of the various segmentation methods

Table 3. Represents the Recall values for various segmentation methods used

Input Image		ROI Segmentation	Watershed	Otsu	Region Growing	Marker Based Watershed	K -means Clustering
1			.262	.025	.278	.786	.771
2			.250	.208	.245	.790	.799
3			.259	.243	.165	.778	.898
4			.260	.215	.176	.876	.884
5			.263	.205	.275	.852	.752
6			.250	.285	.362	.888	.823
7			.251	.294	.365	.747	.896

The Figure 2 and Table 3 represents the graphical and pictorial representation of recall values of various segmentation methods.

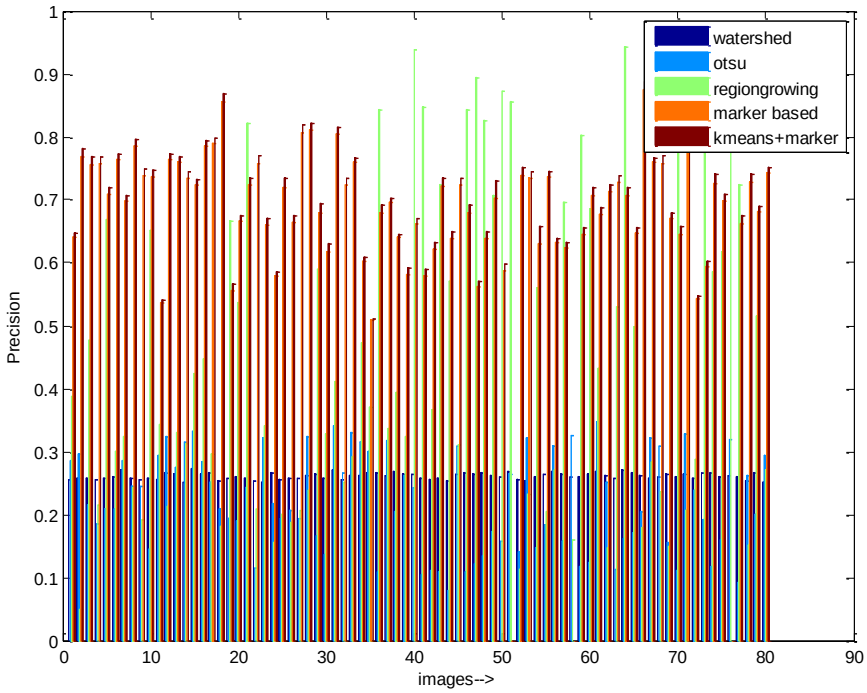

Figure 3. Represents the comparion between the Precision values of the various segmentation method

The Figure 3 represents the graphical precision values for various segmentations used which includes otsu,region growing,marker based watershed segmentation and K Means clustering.

Table 4. Represents the feature 1 and feature 2 comparison

Feature 1	**Feature 2**
3. 40	1. 45
6. 74	2. 90
10. 24	4. 18
13. 50	5. 80
17. 15	7. 21
20. 48	8. 80
23. 98	10. 10

The table 4 clearly shows that the time required for feature 2 is comparatively less than that of feature 1 for the seven number runs as picturized in graph3. Thus, we concluded that feature 2 is much faster than feature 1.

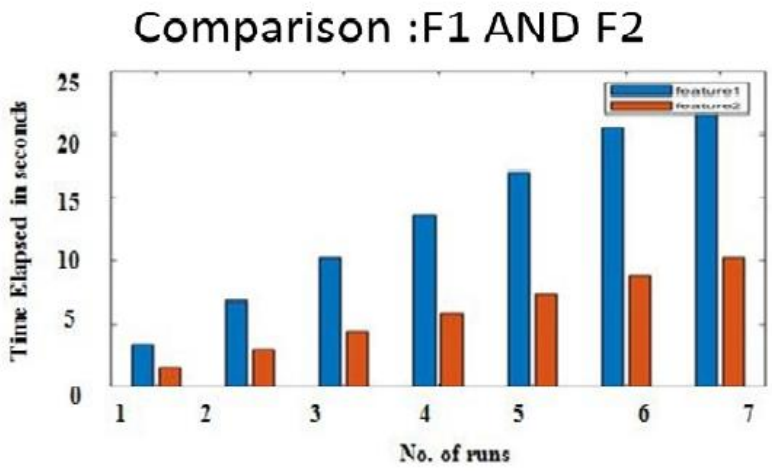

Figure 4. Represents the time elapsed in seconds for both the features with the number of runs

The performance parameter of both the features is being checked out between the number of runs and the time elapsed for the number of runs in the Figure 4. The graphical analyses for both the features, i.e., feature 1 (HOG, LBP, Tamura features, Texture Features, Shape features, and GLCM) and feature 2 (SURF, HOG, PCA), is clearly depicted here. Feature 1 is represented in blue color, whereas feature 2 in red color. The graph is plotted between the number of runs requires for each feature set for the elapsed time duration in seconds.

The performance parameter of a classifier can be analyzed with various parameters like accuracy, precision, recall, and F1 score. The equation for precision and recall has already been discussed in the segmentation session. The remaining terms are accuracy and F1 score.

Accuracy= (Number of correctly classified X-ray chest Images/Total Number of X-ray chest Images) *100 (1)

F1Score = 2* (Precision*Recall)/(Precision-Recall) (2)

$$Recall = TP/(TP+FP) \quad (3)$$

$$Precision=TP/(TP+FN) \quad (4)$$

Table 5. comparison table between different classifiers such as SVM, Neural, Adaboost and LVQ in terms of accuracy, precision ,recall and F1 is made

Classifier	SVM	NEURAL	LVQ	ADABOOST
Precision	**0.98**	**0.89**	**0.96**	**0.92**
Recall	**0.81**	**.81**	**0.95**	**0.99**
Accuracy	**89.47**	**78.89**	**97.25**	**98.52**
F1	**0.89**	**0.81**	**.95**	**0.92**

This analysis has been made for more than 300's of X-ray chest images. The ground truths are obtained from the Tuberculosis Chest X-ray Database (from IEEE dataport).The datasets contains various types of images such as normal and abnormal chest X-ray images. This ground truth images are being preferred for the comparative analysis with the with the proposed hybrid clustering and classification method. The X-ray chest image data set has always provided a better images, so the proposed hybrid methods are initially taken to the normal Chest Web input images and the accuracy, precision, recall and f F1 is being validated.. The trained data set(both positive and negative cases) is fed to the preprocessing, segmentation and feature extraction segment and finally to the classifier model. The suspected input X-ray image undergone the same process as described above is then fed to the classifier section. The classifier model is a combination of both LVQ and Adaboost with the weighted function in a fused segment. The data set (for both normal and abnormal images) has been correctly identified by the AdaBoost algorithm merged with LVQ when compared with the other two. The accuracy, Precision, Recall, and F1score of different classifiers are described in the table 5. The accuracy itself shows that our proposed method yields better result than the other classifiers. That is the merged combination of LVQ and AdaBoost has the highest accuracy, precision, recall and F1 score .It correctly identifies the affected and unaffected chest images (ie normal or abnormal chest X-ray images and will be displayed in the automated approach,so that diagnosis is made faster when compared with other methods)with a percentage of 98.52%,

whereas conventional neural network with an accuracy of 78.94% and SVM with 89.47% accuracy.

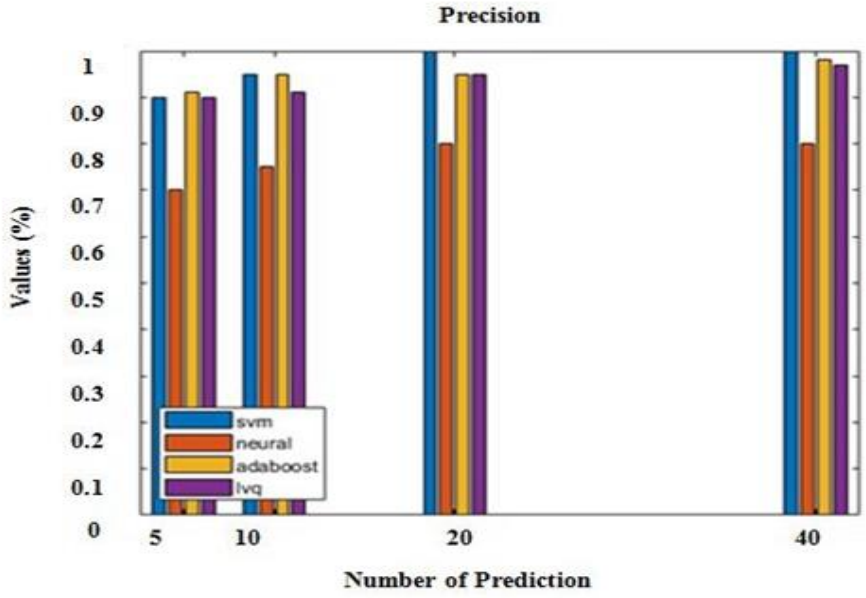

Figure 5. Represents the Precision Values of Adaboost and LVQ compared to SVM and Neural Network

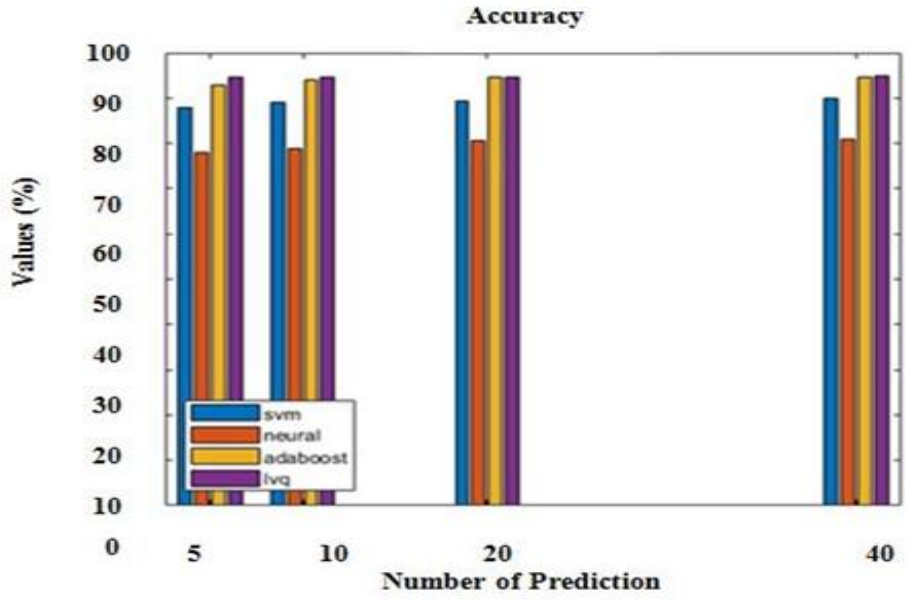

Figure 6. Represents the Accuracy Values of Adaboost and LVQ compared to SVM and Neural Network

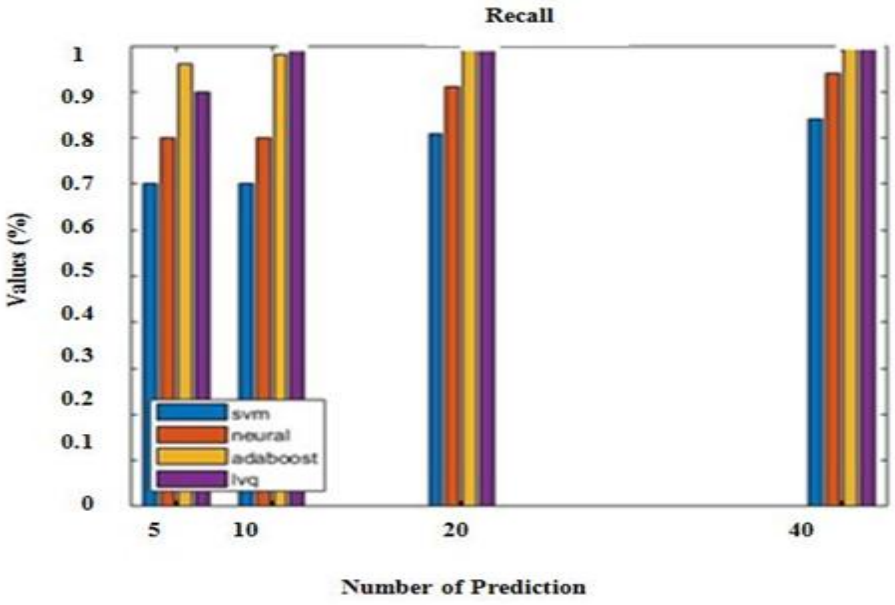

Figure 7. Represents the Recall Values of Adaboost and LVQ compared to SVM and Neural Networks

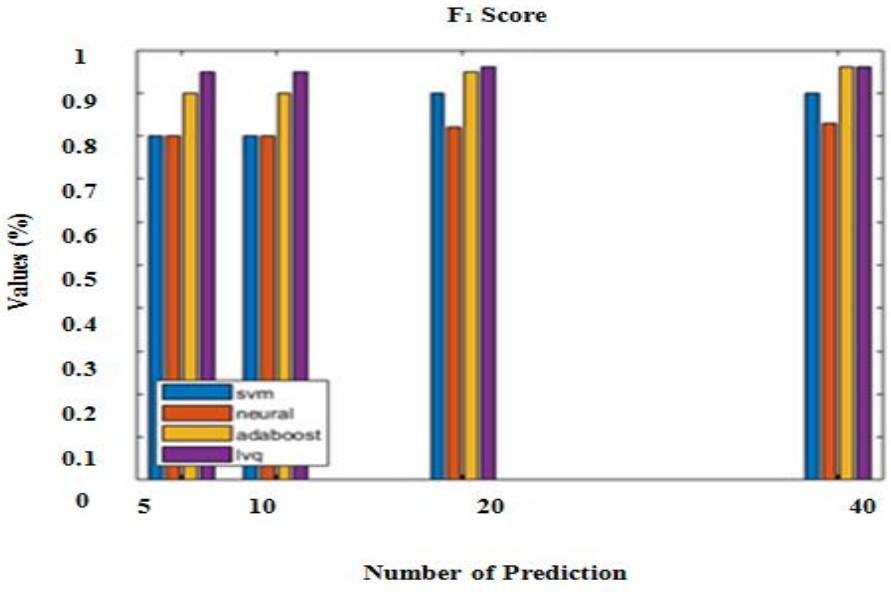

Figure 8. Represents the F1 Score comparison of Adaboost and LVQ with SVM and Neural Network

From the Figures 5,6,7,8- the blue colour in the graph stands for SVM, red for neural network, yellow for AdaBoost, and violet for LVQ. From the above graphs for accuracy, precision, recall, and f1 values of SVM, Neural, Adaboost merged with LVQ, we can observe that the highest value for accuracy, precision, recall, and f1 is secured by the Adaboost algorithm merged with LVQ. This makes the classifier more efficient in analyzing the Tuberculosis of an active person who has been infected with the chest X-ray image, and it will display the output as a normal or abnormal image.

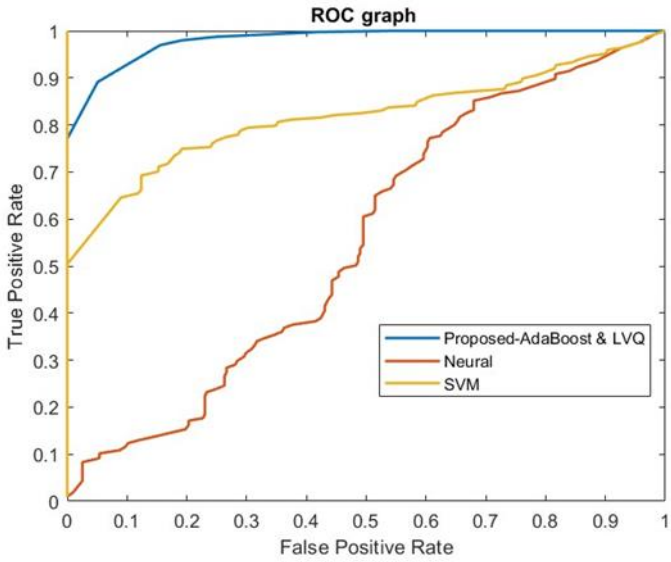

Figure 9. ROC Graph depicting different classifiers

The ROC (Receiver Operating Characteristics)depicts the rate of True Positives with respect to the rate of False Positives in Figure 9.It is the probability curve,which is used to measure the performance of the classifiers.Here from the graph its clearly shown that the proposed-Hybrid combination of Adaboost and LVQ is having higher true positive rate than any other compared classifier methods.The value is much higher when compared to the neural network and SVM(Support Vector Machine).

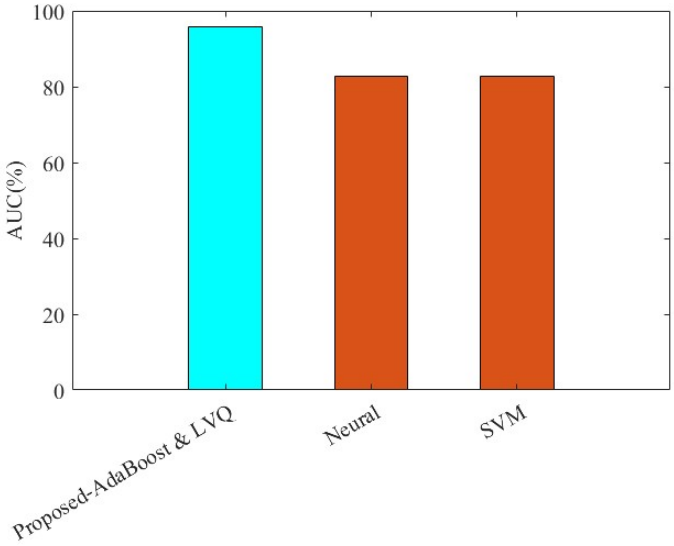

Figure 10. AUC Graph Representation

The AUC Graph in Figure 10 determines the degree or measure of separability.It tells how much the model is capable of distinguishing between classes. Higher the AUC,better the model is at predicting 0 classes at 0 and 1 classes as 1.From the graph it is clearly shown that the proposed classifier excels the other classsifiers interms of AUC.

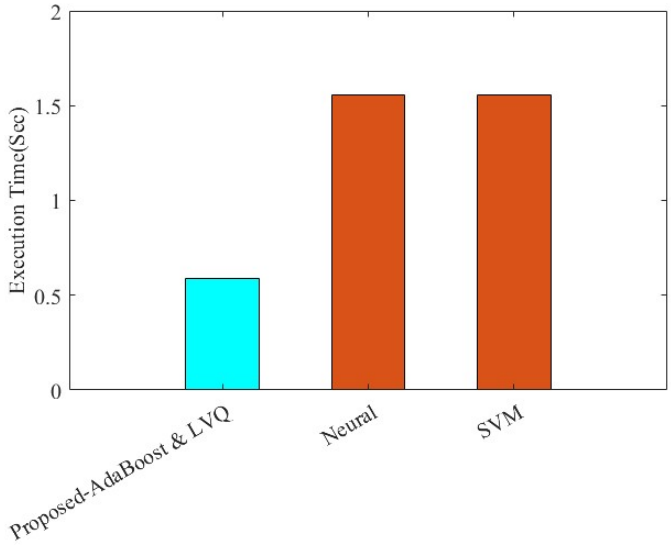

Figure 11. The comparison graph of proposed and other classifiers Execution time

The comparison graph of the Figure 11 depicts the proposed and other classifiers regarding ROC curve, AUC, and execution time is exhibited in graph 5. The ROC curve of the proposed model illustrates higher performance than the SVM and Neural classifiers. Correspondingly, the proposed model's AUC is 95.68, which is higher than the other classifier, whereas the SVM and Neural attain the AUC of 78.34% and 82.79% respectively. Likewise, the execution time of the proposed model is 0.587s, which shows a better performance of the proposed model. So from the results ,it is clear the Hybrid combination of the classifiers excels the rest.

8. CONCLUSION

For rapid as well as accurate analysis of TB detection, a CAD model has been proposed here utilizing a hybrid system with Adaboost and learning vector quantization. This article mainly deals with the segmentation, feature extraction, along with classification of X-ray chest images for the proper diagnosis of TB. The trained data set in the system is then compared with the feature extracted X-ray chest image of the input image to the classifier. Finally, two classifiers (LVQ and Adaboost) are merged into a fusion block that displays the output as normal or abnormal. After that, the experimental analysis is performed in which the performances of the proposed model were analyzed regarding various metrics. The final results show that the proposed model attains the highest accuracy of around 98.52% when compared with other classifier methods. Similarly, the proposed model shows better performance for all other metrics. Hence, it is concluded that the proposed model is highly efficient in the diagnosis of TB. In the future, the work will introduce enhanced preprocessing steps to improve the classification accuracy and also may extend the type classification from abnormal images using novel classification techniques.

REFERENCES

1. Alexandros K, Sameer A, and George T, Segmenting anatomy in chest x-rays for tuberculosis screening, In 2011 *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, (2011), 7779-7782.

2. Wang S. Lu, S.and Zhang Y., A note on the marker-based watershed method for X-ray image segmentation, *Computer Methods and Programs in Biomedicine*, **141** (2017), 1-2.

3. Zhang X., Jiaa F.,Luo S., Liu.G, and Hu.Q, A marker-based watershed method for X-ray image segmentation, *Computer*, **113** (3), (2014) ,894-903.

4. Raviglione M. and Sulis G., Tuberculosis 2015burden, challenges and strategy for control and elimination.*Infect. Dis. Rep.* **8**(2),(2015) 6570.

5. Yang F.,Hamit M., Chaun B. Yan,and Yo.J, Feature Extraction and Classification on Esophagel X-ray Images of Xinjiang Kazak Nationality, *Journal of Health care Engineering*, (2017)

6. Jeyavathana R.B, Balasubramanian R and Pandian A, An Efficient Feature Extraction Method for Tuberculosis detection using Chest Radiographs, *International Journal of Applied Environmental Sciences* **12**(2) (2017) 227-240.

7. Qin, C., Yao, D., Shi, Y., & Song, Z, Computer-aided detection in chest radiography based on artificial intelligence: a survey. *Biomedical engineering online*, **17**(1)(2018), 1-23.

8. Poornimadevi, C. S., & Sulochana, H, Automatic detection of pulmonary tuberculosis using image processing techniques. In *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)* IEEE (2016), 798-802.

9. Cui, W., Chen, S., Yu, T., & Ren, L. Feature extraction of X-ray chest image based on KPCA. In *Proceedings of 2012 2nd International Conference on Computer Science and Network Technology* IEEE (2012) 1263-1266.

10. Vamsha Deepa, N., Krishna, N., & Hemanth Kumar, G, Feature extraction and classification of X-ray lung images using Haralick texture features. In *International*

Conference on Next Generation Computing Technologies , Springer, Singapore(2017).899-907

11. Lingayat, N. S., & Tarambale, M. R,A computer based feature extraction of lung nodule in chest x-ray image. *International Journal of Bioscience, Biochemistry and Bioinformatics*, **3**(6)(2013), 624.

12. Oliveira, L. L. G., e Silva, S. A., Ribeiro, L. H. V., de Oliveira, R. M., Coelho, C. J., & Andrade, A. L. S.,Computer-aided diagnosis in chest radiography for detection of childhood pneumonia. *International journal of medical informatics*, **77**(8)(2008), 555-564.

13. Engin, M., Oğul, H., Ağildere, M., & Sümer, E., An evaluation of image registration methods for chest radiographs. In *2015 SAI Intelligent Systems Conference (IntelliSys)* IEEE (2015)822-827.

14. Horak, K., Klecka, J., Bostik, O., & Davidek, D,Classification of SURF image features by selected machine learning algorithms. In *2017 40th International Conference on Telecommunications and Signal Processing (TSP)* IEEE(2017)636-641.

15. Yu, Y., Hu, P., Lin, J., & Krishnaswamy, P. Multimodal multitask deep learning for X-ray image retrieval. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* Springer(2021), 603-613.

16. Chan, Y. H., Zeng, Y. Z., Wu, H. C., Wu, M. C., & Sun, H. M. Effective pneumothorax detection for chest X-ray images using local binary pattern and support vector machine. *Journal of healthcare engineering*, *2018*.

17. Lu, D., & Weng, Q.A survey of image classification methods and techniques for improving classification performance. *International journal of Remote sensing*, **28**(5), (2007),823-870.

18. Ochs, R. A., Goldin, J. G., Abtin, F., Kim, H. J., Brown, K., Batra, P., ... & Brown, M. S.Automated classification of lung bronchovascular anatomy in CT using AdaBoost. *Medical image analysis*, **11**(3),(2007) 315-324.

19. Melin, P., Amezcua, J., Valdez, F., & Castillo, O. A new neural network model based on the LVQ algorithm for multi-class classification of arrhythmias. *Information sciences*, **279**(2014) 483-497.

20. Ayaz, M., Shaukat, F., & Raja, G. Ensemble learning based automatic detection of tuberculosis in chest X-ray images using hybrid feature descriptors. *Physical and Engineering Sciences in Medicine*, **44**(1)(2021)183-194.

21. Fati, S. M., Senan, E. M., & ElHakim, N. Deep and Hybrid Learning Technique for Early Detection of Tuberculosis Based on X-ray Images Using Feature Fusion. *Applied Sciences*, **12**(14)(2022), 7092.

22. Acharya, B., Acharya, A., Gautam, S., Ghimire, S. P., Mishra, G., Parajuli, N., & Sapkota, B. Advances in diagnosis of Tuberculosis: an update into molecular diagnosis of Mycobacterium tuberculosis. *Molecular biology reports*, **47**(5)(2020), 4065-4075.

23. Eddabra, R., & Ait Benhassou, H. Rapid molecular assays for detection of tuberculosis. *Pneumonia*, **10**(1)(2018), 1-12.

24. Hogeweg, L., Sánchez, C. I., Maduskar, P., Philipsen, R., Story, A., Dawson, R., ... & Van Ginneken, B. Automatic detection of tuberculosis in chest radiographs using a combination of textural, focal, and shape abnormality analysis. *IEEE transactions on medical imaging*, **34**(12)(2015), 2429-2442.

25. Cao, X. F., Li, Y., Xin, H. N., Zhang, H. R., Pai, M., & Gao, L. Application of artificial intelligence in digital chest radiography reading for pulmonary tuberculosis screening. *Chronic Diseases and Translational Medicine*, **7**(01)(2021), 35-40.

LIST OF PUBLICATIONS

1. Paul, E. M., Perumal, B., & Rajasekaran, M. P. Filters used in x-ray chest images for initial stage tuberculosis detection. In *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)* IEEE (2018)235-239

2. Perumal, B., & Paul, E. M. A New Merged Segmentation Technique Used For X-Ray Chest Images. *In 3C Tecnologia. Glosas de innovacion aplicadas a la pyme*.(2021),411-429.

3. Perumal, B., & Paul, E. M. Morphology and Segmentation used in Automated Screening of Tuberculosis. *In Sustainable development in engineering and technology 3ciencians*,(2022),313-328

4. Paul, E. M., & Perumal, B. Computer-Aided Diagnosis using Hybrid Technique for Fastened and Accurate Analysis of Tuberculosis Detection with Adaboost and Learning Vector Quantization. *International Journal of Pattern Recognition and Artificial Intelligence*(2022),2240008.